'Not suited for kids': Language change on speed in Western-European cities

Freek Van de Velde

The decline and fall of the inflectional empire

- A few new synthetic forms notwithstanding (Van Haeringen 1950), Germanic and Romance languages witnessed an increased syntactic complexity at the expense of morphological complexity in their histories.

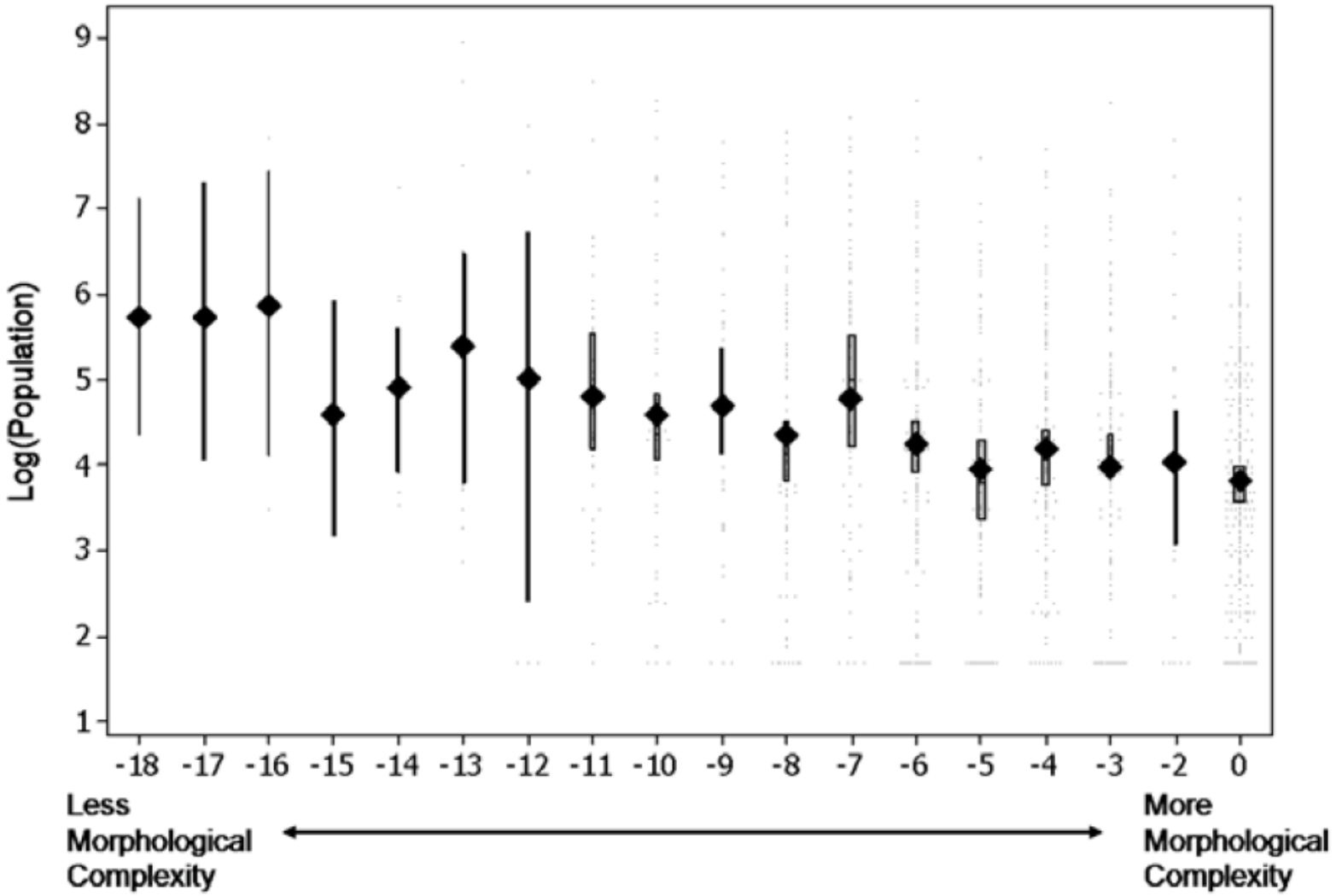
- What happened?

- The adults hijacked the language

Demographic explanations for the analytic-synthetic difference

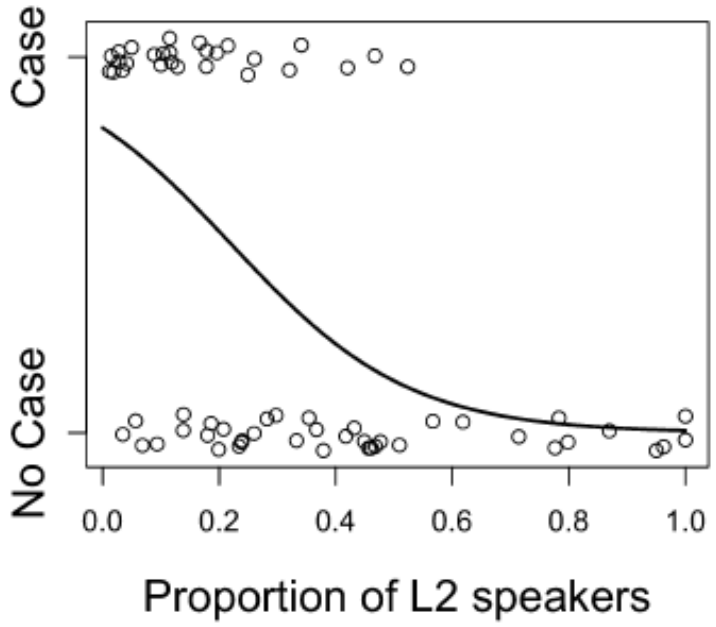
- Lupyan & Dale (2010)
  - Due to evolutionary pressures, languages adapt to their community (see also Christiansen and Chater 2008; Lupyan & Dale 2016)

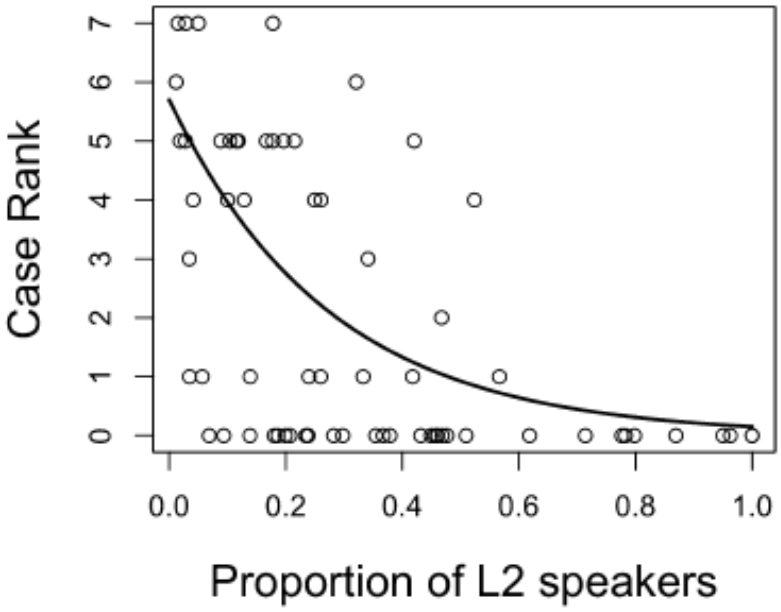
    ⇨ *Linguistic Niche Hypothesis*
  - **Esoteric languages**: morphologic complexity, redundancy, synthetic, favouring L1 acquisition ⇨ smaller languages
  - **Exoteric languages**: analytic-syntactic complexity, transparency, analytic, favouring L2 acquisition ⇨ patterns with bigger languages


Log(Population)
9
8
7
6
5
4
3
2
1
-18
-17
-16
-15
-14
-13
-12
-11
-10
-9
-8
-7
-6
-5
-4
-3
-2
0
Less
Morphological
Complexity
More
Morphological
Complexity

Demographic explanations for the analytic-synthetic difference

- Bentz & Winter (2013)
  - Data:
    - 226 languages using the *SIL Ethnologue*, the *Rosetta project website* and the *UCLA Language Materials Project*; area and family information from *AUTOTYP database,* case information from *WALS*
    - Overlap: 66 languages (26 language families, 16 areas)
  - Operationalisation:
    - L2 speakers: adult L2 speakers as opposed to early bilinguals
    - Case: productive morphological inflections of nouns (loose definition: possessive clitic *-s* in English is counted as case)
  - Method:
    - Generalized linear mixed effects models: logistic regression (case vs. no-case), and negative binomial regression (count of case). Response variable: case; explanatory variable: proportion of L2 speakers
    - Throw in population count in the regression models to see whether it is a predictor on top of the L2 proportion. ⇨ It isn't.


Case
No Case
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of L2 speakers


Case Rank
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of L2 speakers

Demographic explanations for the analytic-synthetic difference

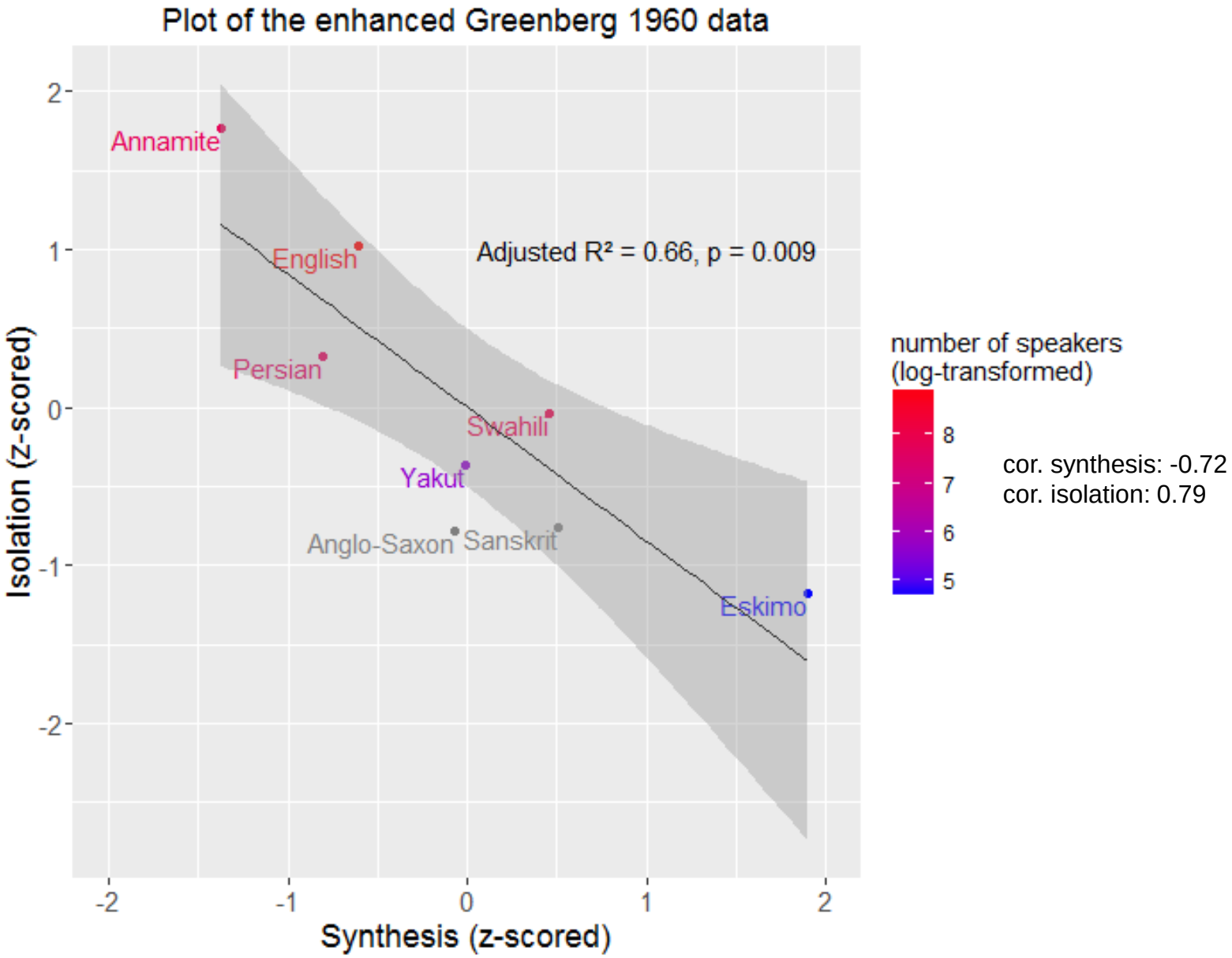
- Lupyan & Dale (2010), Bentz & Winter (2013): synchrony
- What about diachrony? (see Kusters 2003; Szmrecsanyi 2012; Carlier et al. 2012; Haspelmath, forthc.; Haspelmath & Michaelis, forthc.)

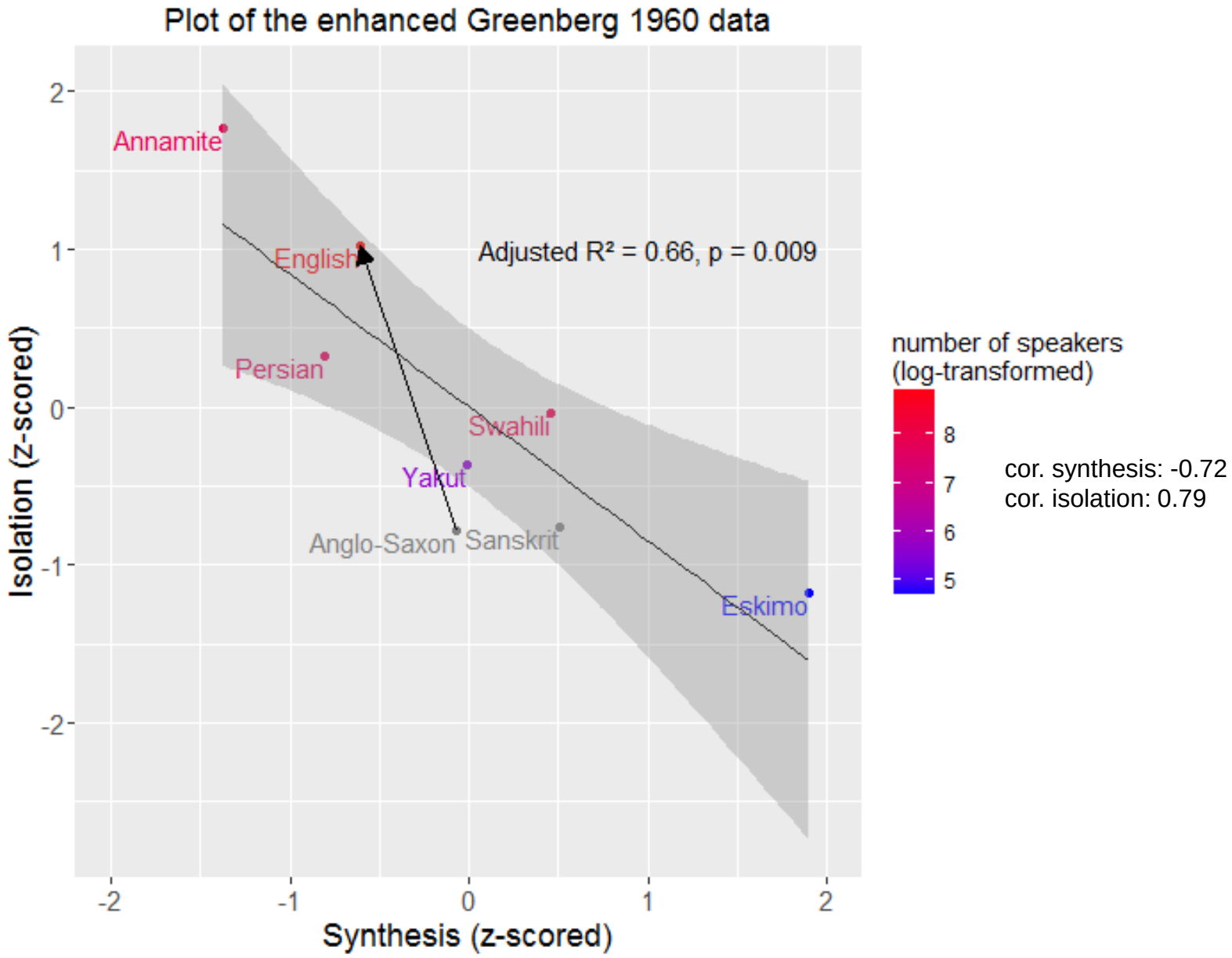
From synchrony to diachrony

Greenberg (1960)

- Index of synthesis (proportion of morphemes to words)
- Index of isolation (proportion of word order as a grammatical marker to the total number of nexus)
- Along with a number of other indices (Index of agglutination, Index of compounding, Index of inflection, Index of prefixation ...).
- Calculated on 100 word stretches of different languages (labour-intensive):

	Sanskrit	Anglo-Saxon	Persian	English	Yakut	Swahili	Annamite	Eskimo
Synthesis	2.59	2.12	1.52	1.68	2.17	2.55	1.06	3.72
Isolation	.16	.15	.52	.75	.29	.40	1.00	.02


Plot of the enhanced Greenberg 1960 data
Annamite
English
Persian
Swahili
Yakut
Anglo-Saxon
Sanskrit
Eskimo
Adjusted R² = 0.66, p = 0.009
Isolation (z-scored)
Synthesis (z-scored)
2
1
0
-1
-2
-2
-1
0
1
2
number of speakers
(log-transformed)
8
7
6
5
cor. synthesis: -0.72
cor. isolation: 0.79


Plot of the enhanced Greenberg 1960 data
Annamite
English
Persian
Swahili
Yakut
Anglo-Saxon
Sanskrit
Eskimo
Adjusted R² = 0.66, p = 0.009
Isolation (z-scored)
Synthesis (z-scored)
2
1
0
-1
-2
-2
-1
0
1
2
number of speakers
(log-transformed)
8
7
6
5
cor. synthesis: -0.72
cor. isolation: 0.79

Changes in Romance and Germanic

Germanic:	English	>	Dutch	>	German	(Van Haeringen 1956)
Romance:	French	>	Italian (N > S)	>	Spanish	(Lamiroy & De Mulder 2011)

- In all core areas of morphosyntax:
  - Gender
  - Pronominal case
  - Tense
  - Aspect
  - Mood
  - Person
  - Articles
  - Raising
  - External possessors (Van de Velde & Lamiroy 2017)
  - ...
- Exceptions: exaptations (Van de Velde & Norde 2016)
- Directly or indirectly related to 'deflection' (Van der Horst 2004: 53)

Demographic correlates

- Can we attribute these changes to demography?
- Historical demography is elusive: no clear data on populations size and migration
- We can work with urbanisation:
  - In pre-industrial times, population growth is too high to be explained solely by natural growth (De Vries 1984:199-266, Howell 2006:208)
  - Migration, leading to koineization (Kerswill 2002), due to an influx of L2 speakers
    - Language diversity was higher in Medieval and Early Modern cities
    - E.g. French in France: 1790: 20% (Certeau et al. 2002), 1893: 75% (Weber 1991: 95)
    - Dialects were often mutually unintelligible

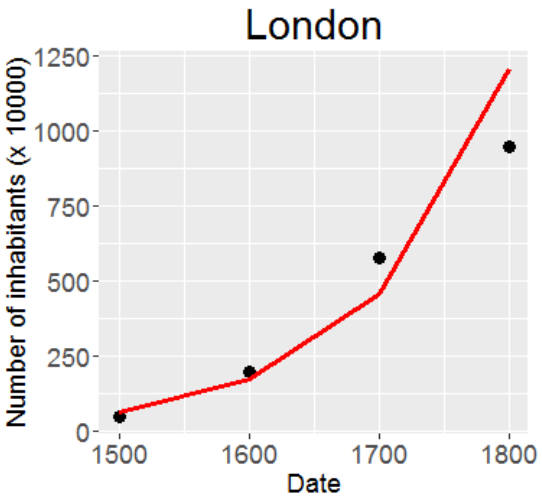
Urbanisation

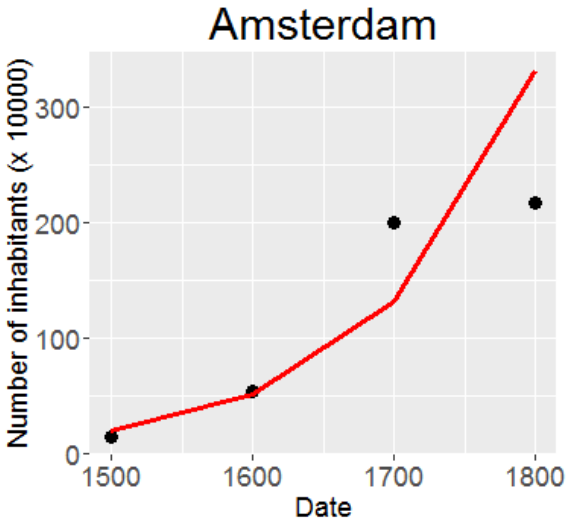
- Do we see more rapid urbanisation in those language areas that analyticise more rapidly?

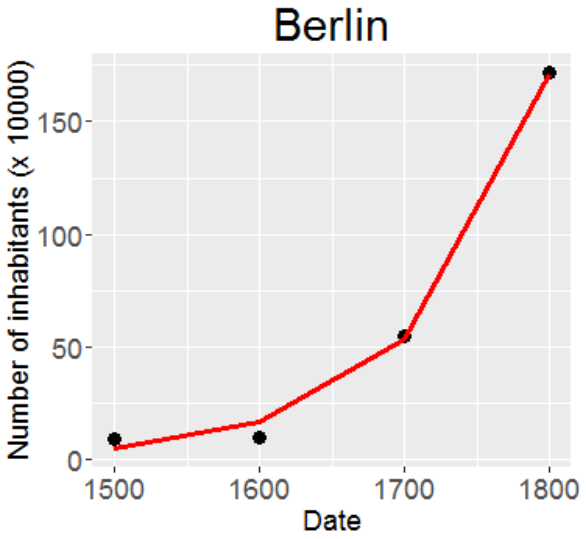
- Between languages
  - English > Dutch > German
  - French > Italian > Spanish

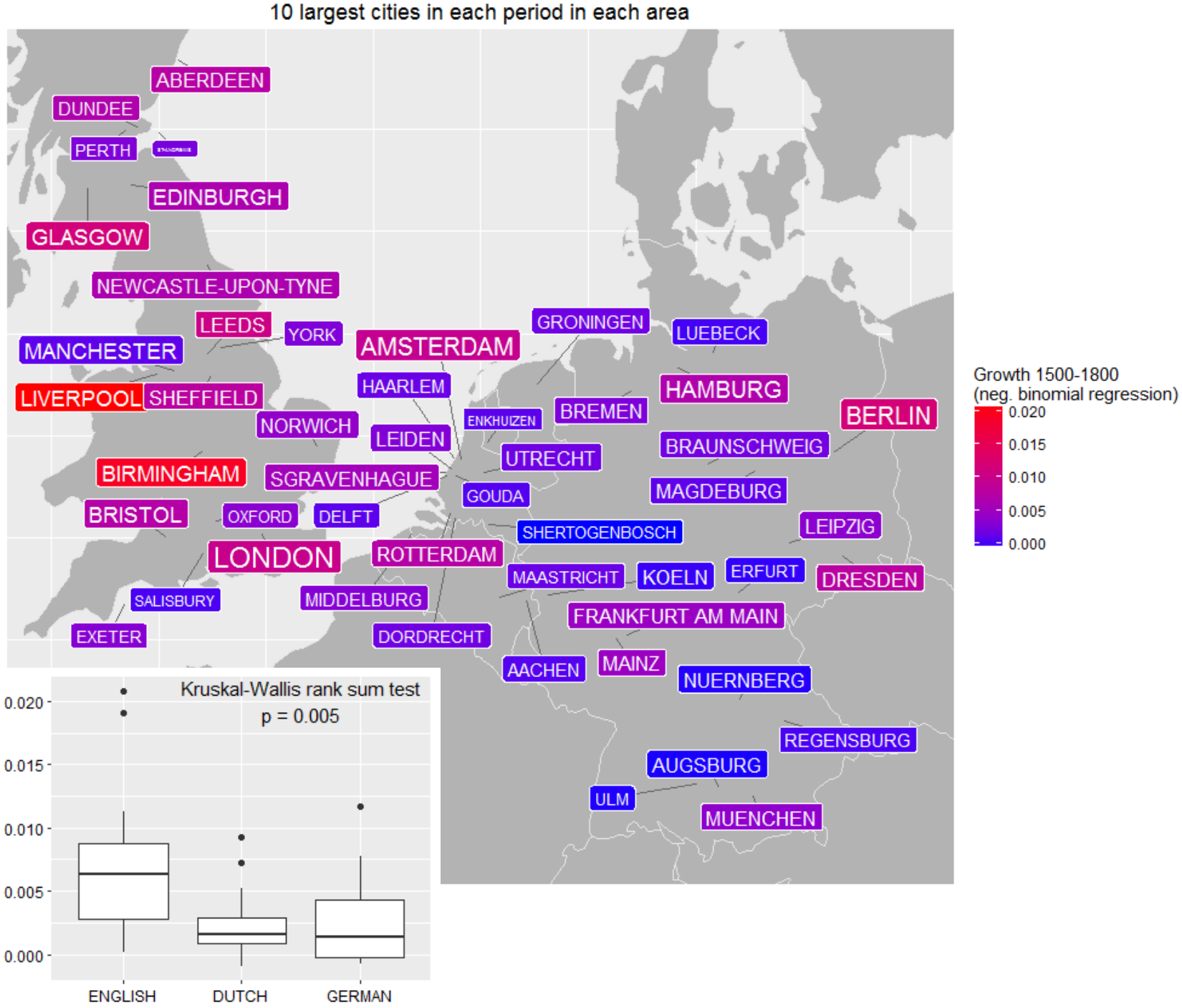
- Within languages
  - Dutch

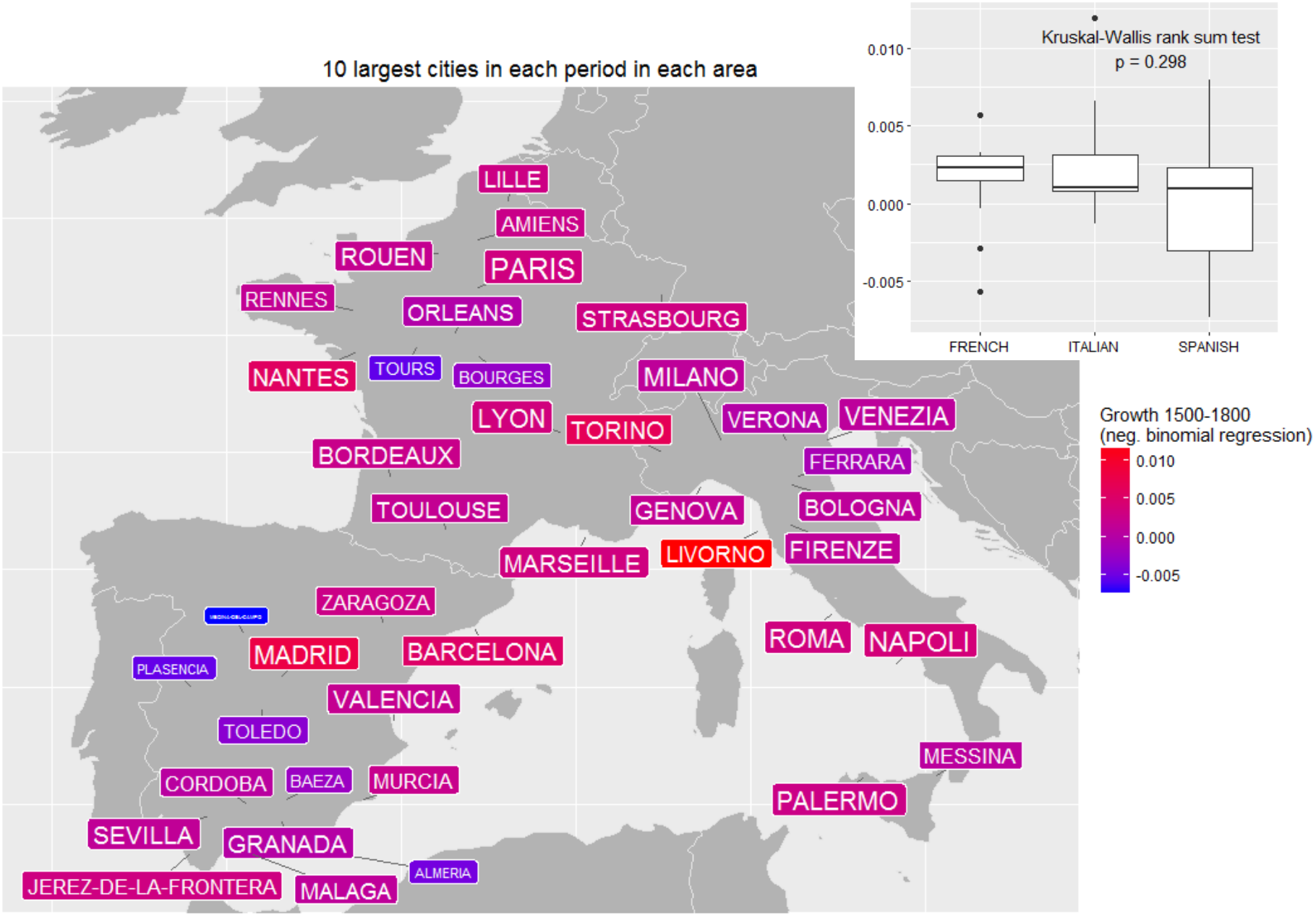
Measuring growth: negative binomial regression

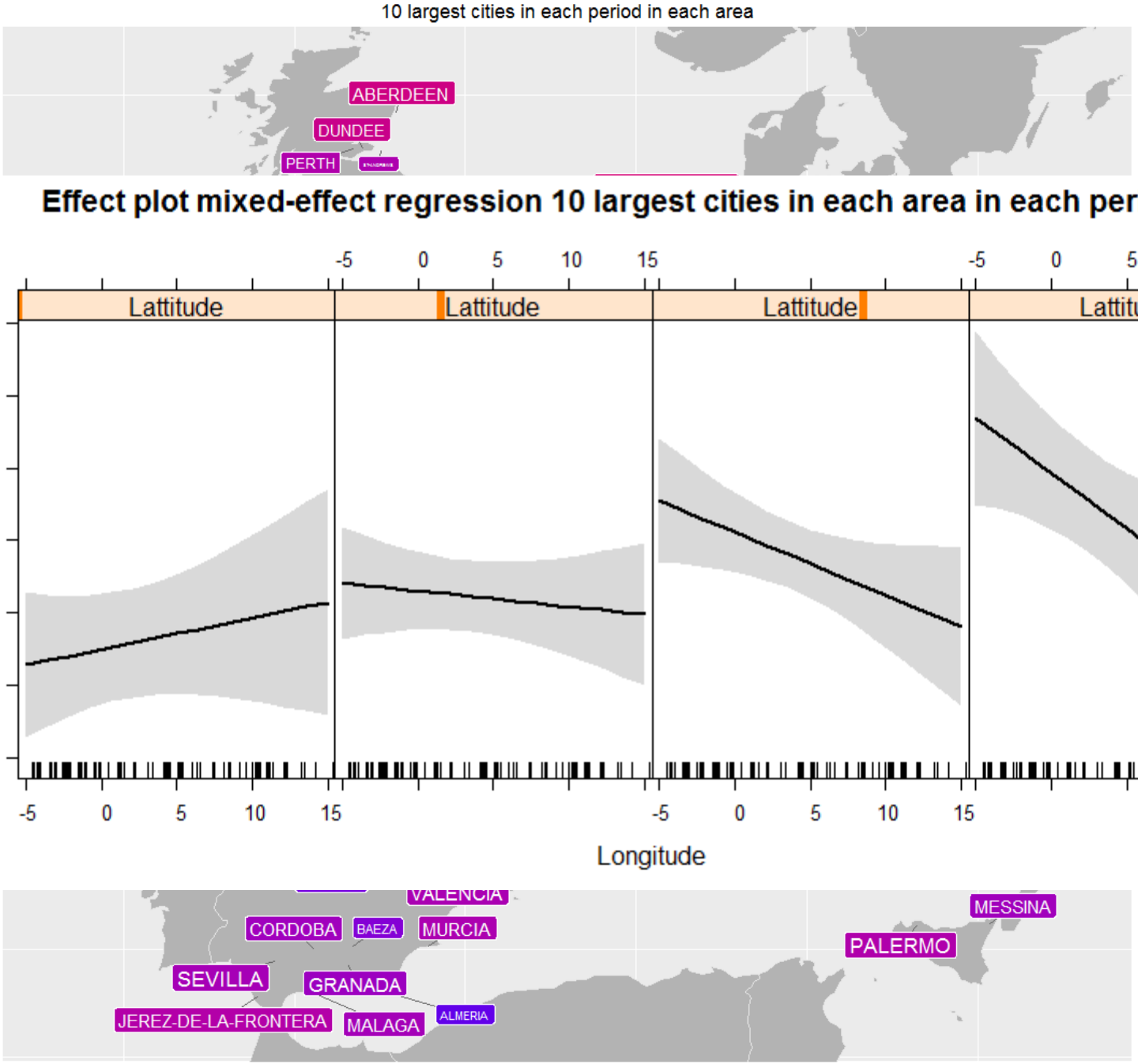


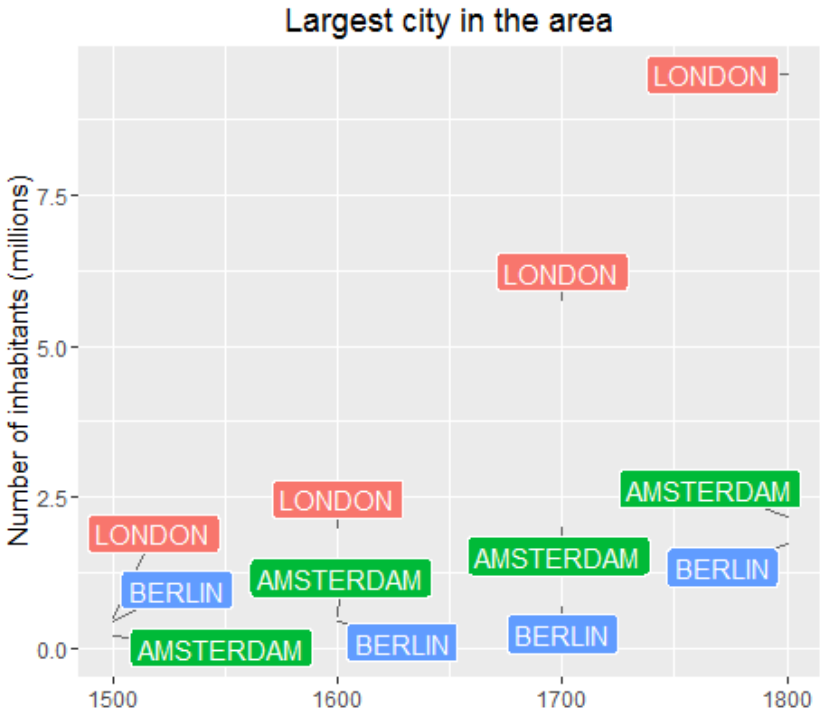


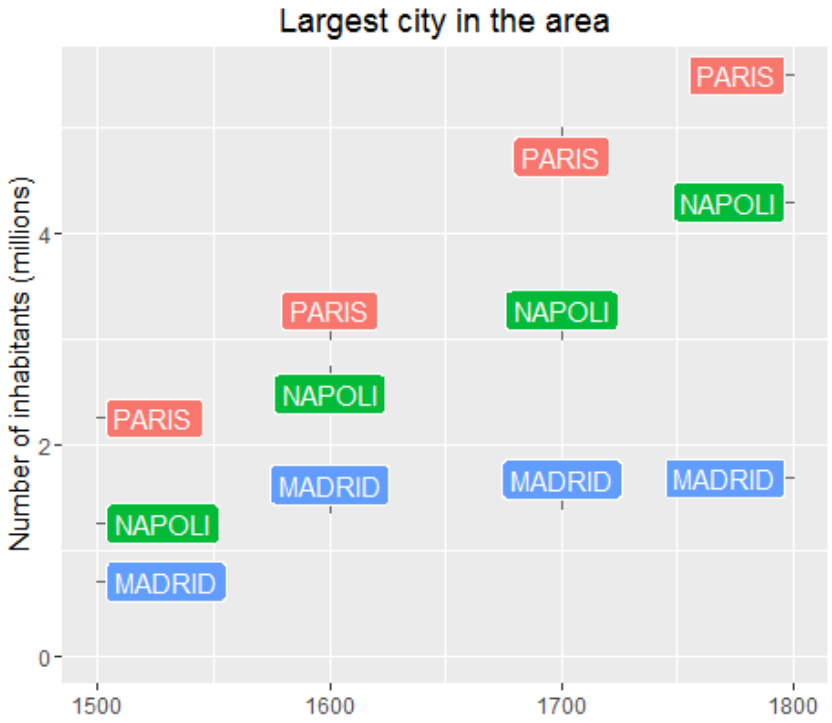



10 largest cities in each period in each area
ABERDEEN
DUNDEE
PERTH
EDINBURGH
GLASGOW
NEWCASTLE-UPON-TYNE
LEEDS
YORK
MANCHESTER
LIVERPOOL
SHEFFIELD
NORWICH
BIRMINGHAM
BRISTOL
OXFORD
DELFT
LONDON
SALISBURY
EXETER
AMSTERDAM
HAARLEM
LEIDEN
SGRAVENHAGUE
GOUDA
ROTTERDAM
MIDDELBURG
DORDRECHT
ENKHUIZEN
UTRECHT
SHERTOGENBOSCH
MAASTRICHT
AACHEN
GRONINGEN
BREMEN
LUEBECK
HAMBURG
BERLIN
BRAUNSCHWEIG
MAGDEBURG
LEIPZIG
KOELN
ERFURT
DRESDEN
FRANKFURT AM MAIN
MAINZ
NUERNBERG
REGENSBURG
AUGSBURG
ULM
MUENCHEN
Growth 1500-1800
(neg. binomial regression)
0.020
0.015
0.010
0.005
0.000
Kruskal-Wallis rank sum test
p = 0.005
0.020
0.015
0.010
0.005
0.000
ENGLISH
DUTCH
GERMAN


10 largest cities in each period in each area
LILLE
AMIENS
ROUEN
PARIS
RENNES
ORLEANS
STRASBOURG
NANTES
TOURS
BOURGES
MILANO
LYON
TORINO
VERONA
VENEZIA
BORDEAUX
FERRARA
TOULOUSE
GENOVA
BOLOGNA
MARSEILLE
LIVORNO
FIRENZE
ZARAGOZA
ROMA
NAPOLI
MADRID
BARCELONA
PLASENCIA
VALENCIA
TOLEDO
MESSINA
CORDOBA
BAEZA
MURCIA
PALERMO
SEVILLA
GRANADA
ALMERIA
JEREZ-DE-LA-FRONTERA
MALAGA
Kruskal-Wallis rank sum test
p = 0.298
0.010
0.005
0.000
-0.005
FRENCH
ITALIAN
SPANISH
Growth 1500-1800
(neg. binomial regression)
0.010
0.005
0.000
-0.005


10 largest cities in each period in each area
ABERDEEN
DUNDEE
PERTH
Effect plot mixed-effect regression 10 largest cities in each area in each period
Lattitude
Lattitude
Lattitude
Lattitude
Growth (negative binomial regression coefficient
0.010
0.008
0.006
0.004
0.002
0.000
-0.002
-5
0
5
10
15
Longitude
VALENCIA
CORDOBA
BAEZA
MURCIA
MESSINA
PALERMO
SEVILLA
GRANADA
JEREZ-DE-LA-FRONTERA
MALAGA
ALMERIA


Largest city in the area
Number of inhabitants (millions)
0.0
2.5
5.0
7.5
1500
1600
1700
1800
LONDON
LONDON
LONDON
AMSTERDAM
AMSTERDAM
AMSTERDAM
BERLIN
BERLIN
BERLIN
BERLIN


Largest city in the area
Number of inhabitants (millions)
0
2
4
1500
1600
1700
1800
PARIS
PARIS
PARIS
PARIS
NAPOLI
NAPOLI
NAPOLI
NAPOLI
MADRID
MADRID
MADRID
MADRID

Case study on simplification: weak preterites

- Germanic languages have two morphological strategies for building preterites (not counting analytic perfects, *he has written a book*):

1. Strong inflection:
   - English *sing – sang*
   - Ablaut, based on Indo-European aspectual system (perfect > preterite)

PIE root *b^h_id^h-	e-grade (present)	o-grade (perfect)
Greek	*peíth-omai*	*$_2$pé-poith-a*
Gothic	*beid-an*	**baid-* (PGm ă < PIE ŏ)

2. Weak inflection
   - English *work – worked*
   - Dental suffix, based on a analytic formation [VERB + *d^heh_1-, *d^hoh_1- ('did')]

Gothic	*beid-an*	**baid-*
Dutch	*beid-en (~ †bijden)*	*beid-de*

Case study on simplification: weak preterites

- Various changes occur:
  - irregularisation (Eng. *buy* – *bought*)
  - one strong ablaut class to another (Du. *heffen* – *hief* < *hoef* (Germ. *hob*, *hub*))
  - weak to strong (Du. *vragen* – *vroeg* (vs. Germ. *fragte*))
  - strong to weak (Eng. *carve* – *carved* < *cearf* (Du. *kerfde* < *karf*))

⇨ Long-term drift, over many centuries

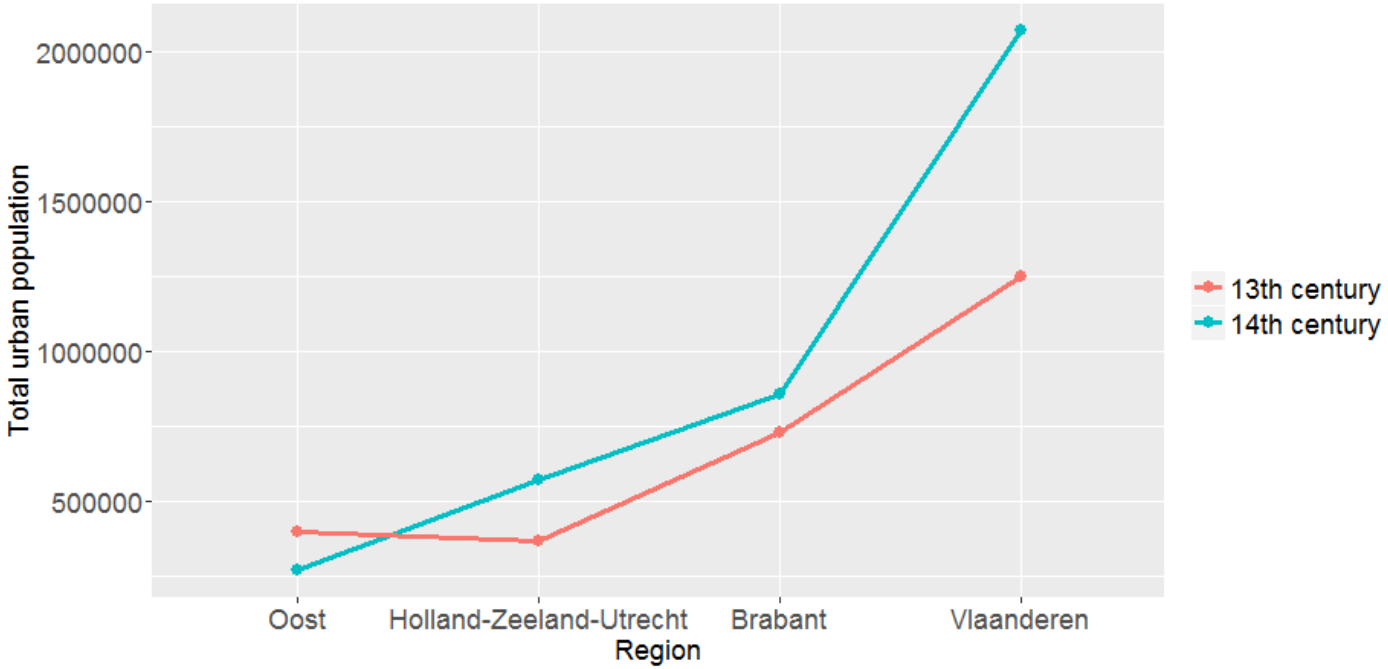
- Morphological simplification ()

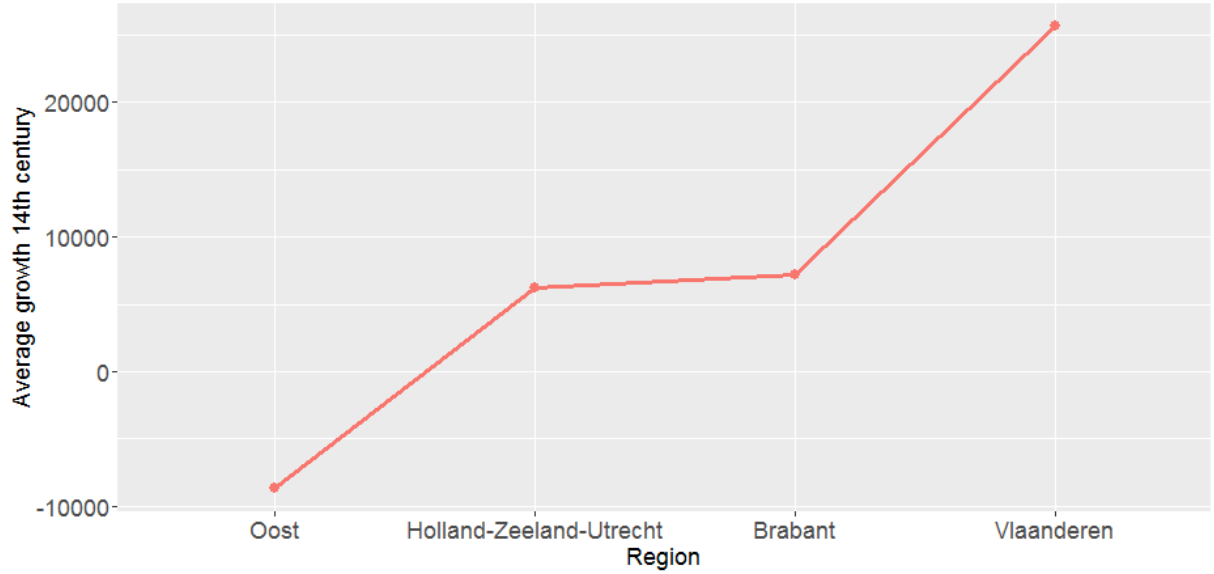
Case study on simplification: weak preterites

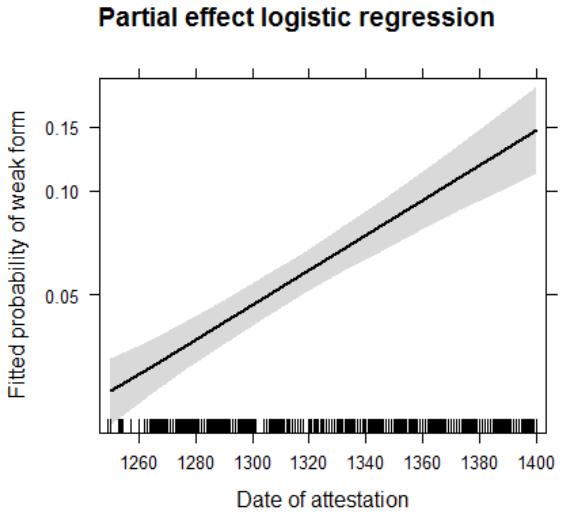
- Various changes occur:
  - irregularisation (Eng. *buy* – *bought*)
  - one strong ablaut class to another (Du. *heffen* – *hief* < *hoef* (Germ. *hob*, *hub*))
  - weak to strong (Du. *vragen* – *vroeg* (vs. Germ. *fragte*))
  - strong to weak (Eng. *carve* – *carved* < *cearf* (Du. *kerfde* < *karf*))

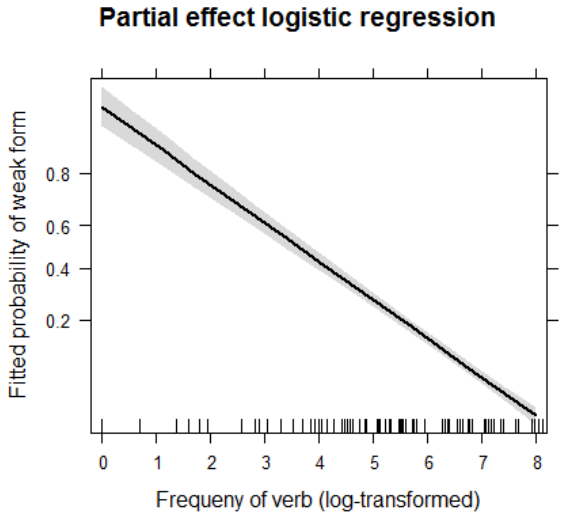
⇨ Long-term drift, over many centuries

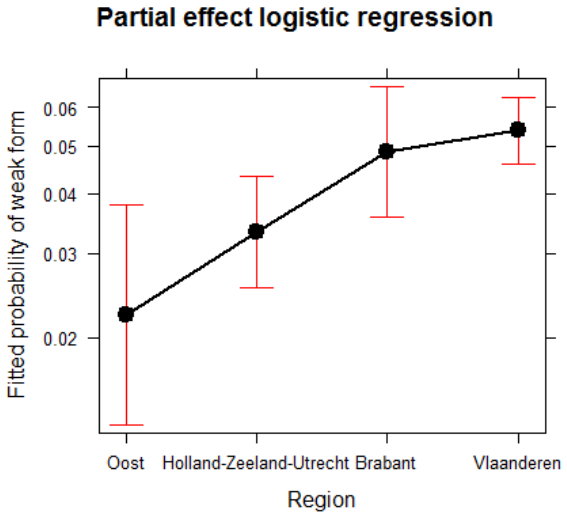
- Follows a English > Dutch > German trend
- But also differentiated within Dutch

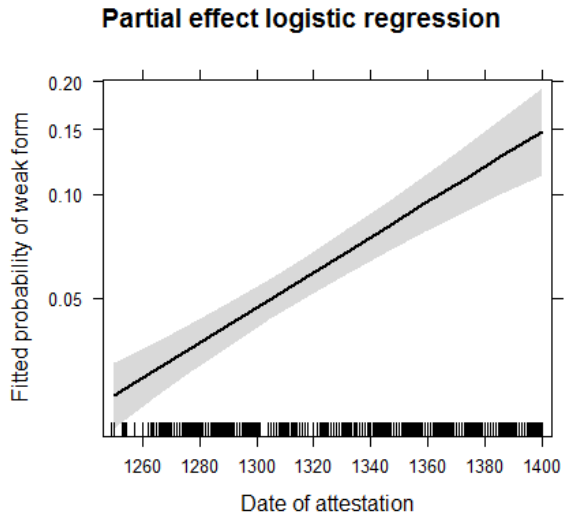

2000000
1500000
1000000
500000
Total urban population
Oost
Holland-Zeeland-Utrecht
Brabant
Vlaanderen
Region
13th century
14th century

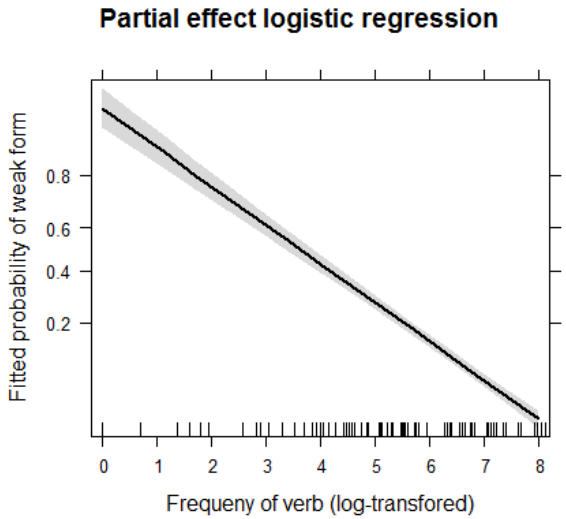

20000
10000
0
-10000
Average growth 14th century
Oost
Holland-Zeeland-Utrecht
Brabant
Vlaanderen
Region

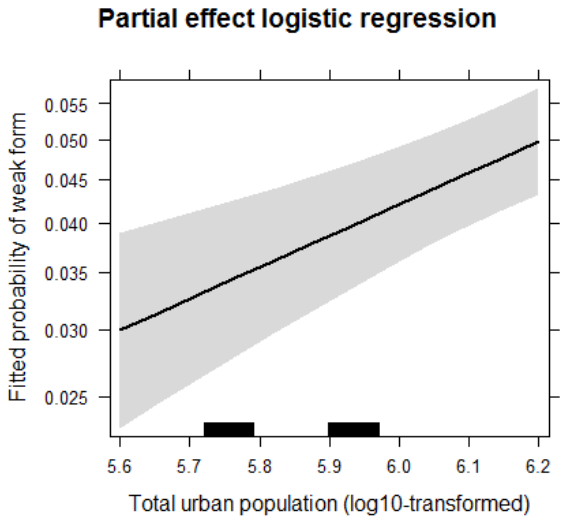

Partial effect logistic regression
Fitted probability of weak form
0.15
0.10
0.05
1260 1280 1300 1320 1340 1360 1380 1400
Date of attestation


Partial effect logistic regression
Fitted probability of weak form
0.8
0.6
0.4
0.2
0 1 2 3 4 5 6 7 8
Frequeny of verb (log-transformed)


Partial effect logistic regression
Fitted probability of weak form
0.06
0.05
0.04
0.03
0.02
Oost
Holland-Zeeland-Utrecht
Brabant
Vlaanderen
Region


Partial effect logistic regression
Fitted probability of weak form
0.20
0.15
0.10
0.05
1260 1280 1300 1320 1340 1360 1380 1400
Date of attestation


Partial effect logistic regression
Fitted probability of weak form
0.8
0.6
0.4
0.2
0 1 2 3 4 5 6 7 8
Frequeny of verb (log-transfored)


Partial effect logistic regression
Fitted probability of weak form
0.055
0.050
0.045
0.040
0.035
0.030
0.025
5.6 5.7 5.8 5.9 6.0 6.1 6.2
Total urban population (log10-transformed)

Conclusions

- Language change is a function of historical demography (L2-speakers)
- between languages of different families
- between languages of the same family (English-Dutch-German)
- within a language (Dutch)

Acknowledgements:

Isabeau De Smet

Dirk Pijpops

Quantitative Lexicology and Variational Linguistics
KU Leuven

Thank you